

Cumulative Conditional Expectation Index

M. Fernández^{*a} and V. A. González-López^b

^aUniversity of Campinas, Brazil

^bDepartment of Statistics, University of Campinas, Brazil

Keywords: Cumulative conditional expectation; Copula distributions.

Abstract

In this paper we study the cumulative conditional expectation function (CCEF) in the copula context. It is shown how to compute CCEF in terms of the cumulative copula function, this natural representation allows to deduce some useful properties, for instance with applications to convex combination of copulas. We introduce approximations of CCEF based on Bernstein polynomial copulas. We introduce estimators for CCEF, which were constructed through Bernstein polynomial estimators for copulas. The estimators are asymptotically normal and biased for CCEF.

1 Introduction

In this paper we explore $\mathbb{E}[V|U \leq u]$ as a function of $u \in (0, 1)$, we denote this quantity as *cumulative conditional expectation function*. One of the motivations of working with this quantity is its use for making decisions in real problems. Our target is to give the foundation for the construction of approximations and estimators for the cumulative conditional expected function. Also, we present in the first

^{*}Corresponding author.

E-mail addresses: mariela@ime.unicamp.br (M. Fernández), veronica@ime.unicamp.br (V. A. González-López).

case the rate of approximation and in the second, the asymptotic distribution of the empirical process related to the estimator.

2 Cumulative conditional expectation

For some applied problems, it is helpful knowing the mean performance of one variable conditioned to another variable restricted to a region and not just to a single observation. In order to reach that knowledge we study the following measure.

Definition 1 *Let (U, V) be a random vector with associated 2-copula C . Then the cumulative conditional expectation function of V on U is given by $R_C(u) = \mathbb{E}[V|U \leq u]$, $\forall u \in (0, 1)$.*

Remark 1 *The cases $u = 1$ and $u = 0$ are excluded because they restrict themselves to the copula regression function. Indeed $\mathbb{E}[V|U \leq 1] = \mathbb{E}[V] = 0.5$ is trivial and $\mathbb{E}[V|U \leq 0] = \mathbb{E}[V|U = 0]$ has been studied in [13].*

We note the previous concept was already explored in [3], but that work focused on some bounds of the measure R_C . The following theorem gives a representation for $R_C(u)$ that let us to simplify its computation.

Theorem 1 *Let (U, V) be a random vector with associated 2-copula C . Then $R_C(u) = 1 - \frac{1}{u} \int_0^1 C(u, v) dv$, $\forall u \in (0, 1)$.*

In order to simplify the notation, hereafter we denote the partial derivatives as done in [5], this is $C^{(1)}(u, v) = \frac{\partial C(u, v)}{\partial u}$, $C^{(2)}(u, v) = \frac{\partial C(u, v)}{\partial v}$, $C^{(1,1)}(u, v) = \frac{\partial^2 C(u, v)}{\partial u^2}$ and $C^{(2,2)}(u, v) = \frac{\partial^2 C(u, v)}{\partial v^2}$. It can be seen that $\mathbb{E}[V|U \leq u]$ is the average of the copula regression function $\mathbb{E}[V|U = u]$.

Corollary 1 *Under the hypotheses of theorem 1, $R_C(u) = \frac{1}{u} \int_0^u \mathbb{E}[V|U = w] dw$, $\forall u \in (0, 1)$.*

Proof. By Theorem 1 we have $R_C(u) = 1 - \frac{1}{u} \int_0^1 C(u, v) dv$, then

$$R_C(u) = \frac{1}{u} \int_0^u dw - \frac{1}{u} \int_0^1 \int_0^u C^{(1)}(w, v) dw dv = \frac{1}{u} \int_0^u \left(1 - \int_0^1 C^{(1)}(w, v) dv\right) dw.$$

The last term is equal to $\frac{1}{u} \int_0^u \mathbb{E}[V|U=w]dw$ according to [13]. ■

By means of Theorem 1 we can compute the cumulative conditional expectation function for several cases, as it was explored in [3]. The following result shows that the cumulative conditional expectation of a convex mixture of copulas is the convex mixture of the cumulative conditional expectations.

Corollary 2 *Under the hypotheses of theorem 1, let C be the convex combination copula, $C(u, v) = \sum_{\gamma=1}^{\Gamma} p_{\gamma} C_{\gamma}(u, v)$ where C_{γ} is a copula function, $0 \leq p_{\gamma} \leq 1$ for $1 \leq \gamma \leq \Gamma$ and $\sum_{\gamma=1}^{\Gamma} p_{\gamma} = 1$. Then, $R_C(u) = \sum_{\gamma=1}^{\Gamma} p_{\gamma} R_{C_{\gamma}}(u)$.*

In the next result we study a specific class of copulas, the polynomial ones. Recalling that a copula has polynomial cross sections in u if it can be written as $C(u, v) = \sum_{i=1}^k \alpha_i(v)u^i$ for each $u \in [0, 1]$, to suitable functions $\alpha_i, i = 1, \dots, k$. More details about these copulas can be found in [10] §3.2. As an immediate consequence of Theorem 1 we have the following corollary.

Corollary 3 *Under the hypotheses of theorem 1, if $C(u, v) = \sum_{i=1}^k \alpha_i(v)u^i$. Then, $R_C(u)$ is a polynomial of degree $k - 1$ given by $R_C(u) = 1 - \sum_{i=0}^{k-1} u^i \int_0^1 \alpha_{i+1}(v)dv$.*

Example 1 *Quadratic and cubic cross sections copulas.*

1. *The Farlie-Gumbel-Morgenstern (FGM) family contains all the copulas with quadratic cross sections in both variables. $C(u, v) = uv + \theta uv(1 - u)(1 - v)$ with $\theta \in [-1, 1]$, then $R_C(u) = (3 + \theta(u - 1))/6$.*
2. *The Lin's iterated FGM family is defined by $C(u, v) = uv + \theta uv(1 - u)(1 - v)(1 + \varphi(1 - u)(1 - v))$ with θ and φ real numbers satisfying $\theta \in [-1, 1]$ and $-1 - \theta \leq \theta(1 + \varphi) \leq (3 - \theta + (9 - 6\theta - 3\theta^2)^{1/2})/2$. C has cubic cross sections in both variables. For $u \in [0, 1]$, $C(u, v) = \sum_{i=1}^3 \alpha_i(v)u^i$ where*

$$\begin{aligned} \alpha_1(v) &= (1 + \theta + \varphi\theta)v + (-\theta - 2\varphi\theta)v^2 + \varphi\theta v^3 \\ \alpha_2(v) &= (-\theta - 2\varphi\theta)v + (\theta + 4\varphi\theta)v^2 - 2\varphi\theta v^3 \\ \alpha_3(v) &= \varphi\theta v - 2\varphi\theta v^2 + \varphi\theta v^3. \end{aligned}$$

Then, $R_C(u) = (\frac{1}{2} - \frac{\theta}{6} - \frac{\varphi\theta}{12}) + (\frac{\theta}{6} + \frac{\theta\varphi}{6})u - \frac{\varphi\theta}{12}u^2$.

3 Bernstein Copulas

Definition 2 Given a 2-copula C and a grid of points $(u_k, v_l) \in [0, 1]^2$ with $k, l = 1, \dots, m$, the Bernstein copula approximation of C of order m is given by

$$B_m C(u, v) = \sum_{k=1}^m \sum_{l=1}^m C\left(\frac{k}{m}, \frac{l}{m}\right) P_{k,m}(u) P_{l,m}(v), \quad \forall (u, v) \in [0, 1]^2,$$

where $P_{j,m}(x) = \binom{m}{j} x^j (1-x)^{m-j}$, $x \in [0, 1]$.

The function $B_m C$ approximates the copula C and is a copula itself. The usefulness of the next theorem is to numerically simplify the computation of R_C , through solving polynomial integrals, when C is a parametric copula described by a complex analytical expression.

Definition 3 Consider a 2-copula C , and a grid of points $(u_k, v_l) \in [0, 1]^2$ with $k, l = 1, \dots, m$. Set the Bernstein approximation of R_C of order m as

$$R_{B_m C}(u) = 1 - \frac{1}{(m+1)u} \sum_{k=1}^m \sum_{l=1}^m C\left(\frac{k}{m}, \frac{l}{m}\right) P_{k,m}(u), \quad \forall u \in (0, 1),$$

where $P_{k,m}(x) = \binom{m}{k} x^k (1-x)^{m-k}$, $x \in [0, 1]$.

Theorem 2 Under the hypotheses of theorem 1, $R_{B_m C}(u)$ converges to $R_C(u)$ pointwise, when $m \rightarrow \infty$.

For appropriate copulas it is possible to give a rate of the approximation for R_C , and also we observe that this rate is improved when u increases.

Theorem 3 Let C be a continuous copula with first order partial derivatives being Lipschitz and $u_0 \in (0, 1)$. For $u \in [u_0, 1)$ it holds $|R_C(u) - R_{B_m C}(u)| \leq \frac{7M}{12u_0^m}$ for a constant M . Then, $R_{B_m C}$ converges uniformly to R_C in $[u_0, 1)$, when $m \rightarrow \infty$.

4 Estimation

Consider a bivariate random sample $\{(U_j, V_j)\}_{j=1}^n$ of the vector (U, V) with associated unknown 2-copula C . Let be R_{U_j} = rank of U_j in $\{U_1, \dots, U_n\}$ and R_{V_j} = rank of V_j in $\{V_1, \dots, V_n\}$. Denote the empirical copula as

$$C_n(u, v) = \frac{1}{n} \sum_{j=1}^n I\left(\frac{R_{U_j}}{n} \leq u\right) I\left(\frac{R_{V_j}}{n} \leq v\right), \quad \forall u, v \in [0, 1].$$

By using the Bernstein estimator of the copula function C proposed in [11] we obtain the estimator of the cumulative conditional expectation function.

Definition 4 *Given a 2-copula C , the estimator of $R_C(u)$ is*

$$\hat{R}_{B_m C_n}(u) = 1 - \frac{1}{(m+1)u} \sum_{k=0}^m \sum_{l=0}^m C_n\left(\frac{k}{m}, \frac{l}{m}\right) P_{k,m}(u), \quad \forall u \in (0, 1),$$

where C_n is the empirical copula.

In order to explore the asymptotic behavior of the estimator $\hat{R}_{B_m C_n}(u)$ we make use of some known results about the Bernstein estimator. In [4] (Theorem 2 and Remark 3) it is proved that for a copula C with bounded third order partial derivatives on $(0, 1)^2$, if $\sqrt{n}/m \rightarrow d$, $0 \leq d < \infty$, when $n \rightarrow \infty$, then the process $\sqrt{n}\left(B_m C_n(u, v) - C(u, v)\right) \rightsquigarrow \mathcal{G}_C(u, v)$ in the space $l^\infty((0, 1)^2)$ of bounded functions, where \rightsquigarrow denotes weak convergence. In addition, the limiting process $\mathcal{G}_C(u, v)$ is a tight Gaussian process with mean function $db(u, v)$ where

$$b(u, v) = \frac{1}{2} \left(u(1-u)C^{(1,1)}(u, v) + v(1-v)C^{(2,2)}(u, v) \right) \quad (1)$$

and covariance function given by $E[h(u, v)h(u', v')]$ for $0 < u, u', v, v' < 1$ with

$$h(u, v) = I(U \leq u, V \leq v) - C(u, v) - C^{(1)}(u, v)(I(U \leq u) - u) - C^{(2)}(u, v)(I(V \leq v) - v). \quad (2)$$

This time we consider the process $\frac{1}{u} \int_0^1 \mathcal{G}_C(u, v) dv$, with $u \in (0, 1)$, which is the process that defines the asymptotic behavior of the proposed estimator of R_C .

Theorem 4 *Let C be a 2-copula with bounded third order partial derivatives on $(0, 1)^2$. If n and m are positive integers such that $\sqrt{n}/m \rightarrow d$ with $0 \leq d < \infty$ when $n, m \rightarrow \infty$, then for $u \in (0, 1)$,*

$$\sqrt{n} \left(\hat{R}_{B_m C_n}(u) - R_C(u) \right) \rightsquigarrow -\frac{1}{u} \int_0^1 \mathcal{G}_C(u, v) dv.$$

$-\frac{1}{u} \int_0^1 \mathcal{G}_C(u, v) dv$ is a Gaussian process with mean function $d(\frac{1}{2} - R_C(u)) + \frac{d(u-1)}{2} \int_0^1 C^{(1,1)}(u, v) dv$ and variance function

$$\begin{aligned} \text{Var} \left[-\frac{1}{u} \int_0^1 \mathcal{G}_C(u, v) dv \right] &= -4R_C^2(u) + \left(4 - \frac{1}{u}\right) R_C(u) - 4\left(2 - \frac{1}{u}\right) R_C(u) \mathcal{H}_1(u, u) \\ &\quad + 2\left(3 - \frac{2}{u}\right) \mathcal{H}_1(u, u) - 4\left(1 - \frac{1}{u}\right) \mathcal{H}_1^2(u, u) - 2 + \frac{1}{u} + \frac{1}{u^2} \int_0^1 (C^{(2)}(u, v))^2 v dv. \end{aligned}$$

where $2\mathcal{H}_1(u, u) = \int_0^1 C^{(1)}(u, v) dv$.

Remark 2 *The covariance function of the process $-\frac{1}{u} \int_0^1 \mathcal{G}_C(u, v) dv$ is*

$$\begin{aligned} \text{Cov} \left[-\frac{1}{u} \int_0^1 \mathcal{G}_C(u, v) dv, -\frac{1}{u'} \int_0^1 \mathcal{G}_C(u', v') dv' \right] &= \\ &= \frac{1}{uu'} \int_0^1 \int_0^1 C(u \wedge u', v \wedge v') dv' dv + (1 - R_C(u))(R_C(u') - 1) + \mathcal{H}_3(u, u') + \mathcal{H}_3(u', u) \\ &\quad + \left(\frac{1}{u \vee u'} - 1\right) \int_0^1 C^{(1)}(u', v) dv \int_0^1 C^{(1)}(u, v) dv - \mathcal{H}_2(u, u') - \mathcal{H}_2(u', u) \\ &\quad + \frac{1}{uu'} \int_0^1 \int_0^1 C^{(2)}(u', v') C^{(2)}(u, v) (v \wedge v') dv' dv - R_C(u) R_C(u') + \mathcal{H}_1(u, u') + \mathcal{H}_1(u', u). \end{aligned}$$

with

$$\begin{aligned}
\mathcal{H}_1(w_1, w_2) &= \frac{1}{w_1 w_2} \int_0^1 C^{(1)}(w_1, v) dv \int_0^1 C(w_1, v) C^{(2)}(w_2, v) dv \\
\mathcal{H}_2(w_1, w_2) &= \frac{1}{w_1 w_2} \int_0^1 \int_0^1 C^{(2)}(w_2, v) C(w_1, v \wedge v') dv dv' - (1 - R_C(w_1)) R_C(w_2) \\
\mathcal{H}_3(w_1, w_2) &= \left(\frac{1}{w_1 \vee w_2} (R_C(w_1 \wedge w_2) - 1) + 1 - 2R_C(w_1) \right) \int_0^1 C^{(1)}(w_2, v) dv.
\end{aligned}$$

for $w_1 \vee w_2 = \max\{w_1, w_2\}$ and $w_1 \wedge w_2 = \min\{w_1, w_2\}$.

References

- [1] Fernández, M. and González-López, V. A. (2013). A Bayesian approach for convex combination of two Gumbel-Barnett copulas. *AIP Conference Proceedings*, **1558**, 1491-1494. Doi: 10.1063/1.4825802.
- [2] Fernández, M. and González-López, V. A. (2013). A copula model to analyze minimum admission scores. *AIP Conference Proceedings*, **1558**, 1479-1482. Doi: 10.1063/1.4825799.
- [3] Fernández, M. and González-López, V. A. (2015). Bounds for the cumulative conditional expectation function. *AIP Conference Proceedings*, **1648**, 060002. Doi: 10.1063/1.4912372.
- [4] Janssen, P., Swanepoel, J. and Veraverbeke, N. (2012). Large sample behavior of the Bernstein copula estimator. *Journal of Statistical Planning and Inference*, **142**, 1189-1197. Doi: 10.1016/j.jspi.2011.11.020.
- [5] Janssen, P., Swanepoel, J. and Veraverbeke, N. (2014). A note on the asymptotic behavior of the Bernstein estimator of the copula density. *Journal of Multivariate Analysis*, **124**, 480-487. Doi: 10.1016/j.jmva.2013.10.009.
- [6] Hu, L. (2006). Dependence patterns across financial markets: a mixed copula approach. *Applied Financial Economics*, **16**, 717-729. Doi: 10.1080/09603100500426515.

- [7] Kim, D. and Kim, J.-M. (2013). Analysis of directional dependence using asymmetric copula-based regression models. *Journal of Statistical Computation and Simulation*, **84** (9), 1990-2010. Doi: 10.1080/00949655.2013.779696.
- [8] Li, X., Mikusiński, P. and Taylor, M. D. (1998). Strong approximations of copulas. *Journal of Mathematical Analysis and Applications* **225** 608-623. Doi:10.1006/jmaa.1998.6056.
- [9] Muddapur, M. V. (2008). A Remark on the Paper A Note on Directional Dependence in Regression Setting By Engin A. Sungur, *Communications in Statistics - Theory and Methods*, **37** (3), 386-387. Doi: 10.1080/03610920701441391.
- [10] Nelsen, R. (2006). *An Introduction to Copulas*, Lecture Notes in Statistics 139. New York, USA: Springer-Verlag.
- [11] Sancetta, A. and Satchell, S. (2004). The Bernstein copula and its applications to modeling and approximations of multivariate distributions. *Econometric Theory* **20**, 535-562. Doi:10.1017/S026646660420305X.
- [12] Sungur, E. A. (2005). A note on directional dependence in regression setting. *Communications in Statistics - Theory and Methods*, **34** (9-10), 1957-1965. Doi: 10.1080/03610920500201228.
- [13] Sungur, E. A. (2005). Some Observations on Copula Regression Functions. *Communications in Statistics - Theory and Methods*, **34** (9-10), 1967-1978. Doi: 10.1080/03610920500201244.
- [14] Van der Vaart, A. W. (2000). *Asymptotic Statistics*. Cambridge University Press.
- [15] Vrac M., Billard L., Diday E., Chédin A. (2012). Copula analysis of mixture models. *Computational Statistics*, **27**, 427-457. Doi: 10.1007/s00180-011-0266-0.